Research Paper ■

Comparison of Information Content of Structured and Narrative Text Data Sources on the Example of Medication Intensification

ALEXANDER TURCHIN, MD, MS, MARIA SHUBINA, DSC, EUGENE BREYDO, PHD, MERRI L. PENDERGRASS, MD, PHD, JONATHAN S. EINBINDER, MD, MPH

A b s t r a c t   **Objective:** To compare information obtained from narrative and structured electronic sources using anti-hypertensive medication intensification as an example clinical issue of interest.

Design: A retrospective cohort study of 5,634 hypertensive patients with diabetes from 2000 to 2005.

Measurements: The authors determined the fraction of medication intensification events documented in both narrative and structured data in the electronic medical record. The authors analyzed the relationship between provider characteristics and concordance between intensifications in narrative and structured data. As there is no gold standard data source for medication information, the authors clinically validated medication intensification information by assessing the relationship between documented medication intensification and the patients' blood pressure in univariate and multivariate models.

Results: Overall, 5,627 (30.9%) of 18,185 medication intensification events were documented in both sources. For a medication intensification event documented in narrative notes the probability of a concordant entry in structured records increased by 11% for each study year ($p < 0.0001$) and decreased by 19% for each decade of provider age ($p = 0.035$). In a multivariate model that adjusted for patient demographics and intraphysician correlations, an increase of one medication intensification per month documented in either narrative or structured data were associated with a 5–8 mm Hg monthly decrease in systolic and 1.5–4 mm Hg decrease in diastolic blood pressure ($p < 0.0001$ for all).

Conclusion: Narrative and structured electronic data sources provide complementary information on anti-hypertensive medication intensification. Clinical validity of information in both sources was demonstrated by correlation with changes in blood pressure.

■ **J Am Med Inform Assoc.** 2009;16:362–370. DOI 10.1197/jamia.M2777.

Introduction

A large fraction of medical data are contained in narrative documents.[1] As electronic medical record (EMR) systems grow more prevalent,[2] narrative information is increasingly being entered in digital format and thus becomes amenable to computational extraction. Since the late 1990's, a large number of tools have been successfully developed for this purpose.[3–9]

Electronic medical record (EMR) systems employ increasingly rich data models that offer a large variety of options for structured data entry.[10] Data available from the EMR systems frequently includes electronic prescribing information problem and allergy lists, structured note templates, inpatient and outpatient orders, and laboratory results, among others. These data sets have great potential for use in clinical research and/or quality of care surveillance.[11,12]

Not surprisingly, the information in the narrative and structured data sources in the EMR frequently overlap. Physicians typically document all facts pertinent to patient care in narrative notes; at the same time, many of these facts are also entered into the structured data fields in the EMR. It is not known how the data from narrative and structured electronic information compare, to what extent they are overlapping or complementary, and which one better represents reality.

Background

Elevated blood pressure is the most common treatable cardiovascular risk factor[13] and is one of the major risk factors for macro- and micro-vascular complications in patients with diabetes.[14–16] Nevertheless, a majority of diabetic patients with hypertension do not have their blood pressure under control.[17,18] The reasons for poor blood pressure control are not completely understood but lack of appropriate intensification of anti-hypertensive medications is thought

to be a significant contributing factor.[19] Rate of treatment intensification, when faced with an abnormal finding (e.g., elevated blood pressure or blood glucose), is an emerging measure of quality of care[20,21] which has been endorsed as "tightly linked" to clinical outcomes.[22] Several studies have shown that intensification of anti-hypertensive medications is associated with greater decrease in blood pressure and higher degree of blood pressure control.[19,23]

Most studies of treatment intensification in care of hypertension to date relied on data manually extracted from medical records—a labor-intensive and expensive process. With the advent of EMR systems it has become possible to use both structured medication data from electronic prescribing systems and medication information obtained from digitized narrative documents such as physician notes. Both sources of information have potential advantages and disadvantages. For example, while physicians usually document all medication changes in their notes, the accuracy of extracting this information computationally will likely be less than 100%. By contrast, though electronic medication records may be easier to analyze, not all changes in medication may be documented there and some changes in prescription may not reflect what the patient is actually taking. The balance of these potential pros and cons has not been well studied and it is not known whether one or both sources of information should be used.

The outpatient EMR used at our institution allows clinicians to enter both structured electronic prescriptions and narrative documents, such as progress notes. We have previously developed and validated a high-fidelity semantic text processor that identifies documentation of blood pressure measurements and anti-hypertensive medication intensification in the text of physician notes.[24] In this study we have compared medication intensification data obtained from the text of the notes and from the structured EMR medication records of hypertensive patients with diabetes. We assessed clinical validity of both data sources by evaluating their relationship to the patients' blood pressure.

Research Aims

The specific goals of the study were as follows:

1. To determine the concordance of anti-hypertensive medication intensification information between the narrative physician notes and structured medication lists in the EMR
2. To identify factors that affect concordance between narrative and structured medication intensification data
3. To assess clinical validity of narrative physician notes and structured medication lists in the EMR as the sources of anti-hypertensive medication intensification information.

Methods

Design

Concordance of Medication Intensification Information in Narrative and Structured EMR Data
We carried out a retrospective analysis of EMR data to determine the fraction of all anti-hypertensive medication intensification events documented in either narrative notes or structured medication lists that were shared between the two sources (primary outcome variable). A single medica-

tion intensification event documented in either narrative or structured data served as the unit of analysis.

Factors That Affect Concordance of Narrative and Structured Medication Data
It is not known how provider characteristics affect patterns of medication documentation in the EMR. We analyzed the relationship between the presence of a structured medication intensification record corresponding to a particular medication intensification record in narrative notes (binary primary outcome variable) and the following predictor variables: (1) provider age; (2) provider gender; (3) study year; and (4) the number of structured medication records entered by the provider prior to the date of the narrative note (representing provider's experience with structured EMR medication records). A single medication intensification event documented in narrative notes served as the unit of analysis.

Clinical Validity of Medication Intensification Data in Narrative and Structured EMR Records
Anti-hypertensive medication intensification is known to be associated with lower blood pressure.[19,23] We therefore evaluated the relationship of frequency of medication intensification documented in (a) narrative notes only; (b) structured EMR records only; and (c) both data sources (predictor variables) and the average monthly change in systolic and diastolic blood pressure (primary outcome variables). A single patient served as the unit of analysis.

Data Sources
Partners Healthcare System is comprised of several academic and community hospitals and private physician groups in eastern Massachusetts, including the founding members Brigham and Women's Hospital and Massachusetts General Hospital. Most physicians affiliated with these two hospitals use an internally developed outpatient electronic medical record (EMR) system Longitudinal Medical Record (LMR).[25] The LMR system allows for entry of both structured dictionary-based data (e.g., medications, allergies, problems) as well as narrative text (e.g., progress notes, radiology and pathology reports, and others). For this study we compared anti-hypertensive medication intensification information obtained from the structured medication entries and through analysis of the text of narrative physician notes in the LMR. Physician age and gender were obtained from the public records of the Massachusetts Board of Registration in Medicine. All data were deidentified.

Patients
We included in our study all patients who were followed in primary care practices at either Brigham and Women's Hospital or Massachusetts General Hospital for at least two years between 1/1/00 and 8/31/05, were at least 18 years old, had a documented diagnosis of diabetes mellitus, and had at least one encounter during the study period where blood pressure above the target level was recorded. Patients who had at least one encounter with an endocrinologist during the study period that addressed diabetes (as ascertained using billing data and computerized analysis of the text of the notes) were excluded. These selection criteria are similar to the ones used in previously published studies of treatment intensification in patients with hypertension.[19,26]

Diagnosis of diabetes was ascertained by analyzing the text of physician notes in the electronic medical record as previously reported.[27] We used 129 and 84 mm Hg as the recommended treatment goals of systolic and diastolic blood pressure, in accordance with the guidelines published before the beginning of the study period.[28] Patients whose only recorded elevated blood pressure measurement was during the last study encounter were excluded. Patients who never had blood pressure above the treatment target for more than a month were excluded.

Study Measurements

Treatment intensification was defined as an initiation of a new or an increase in the dose of an existing anti-hypertensive medication.[23,26,29] We classified a change from one anti-hypertensive medication to another as treatment intensification because no validated means of comparing dose strengths between different anti-hypertensive drugs exists.

Blood pressure values and anti-hypertensive treatment intensification were computationally abstracted from the text of physician notes in the electronic medical record using a previously validated text processor. The sensitivity and specificity of this method are 91 and 96%, respectively, for identification of blood pressure values, and 84 and 95% for identification of anti-hypertensive treatment intensification.[24] Treatment intensifications documented in the structured medication entries in the EMR were identified by analyzing a database that contains all changes made to each structured medication record.

Medication intensifications in physician notes and EMR records that had the same *service date* were treated as identical for the purpose of data analysis because the name of the medication being intensified could not always be ascertained computationally from the narrative notes. *Service date* in our institution's EMR indicates the date of the provider-patient encounter to which the electronic transaction pertains and may be different from the date when the record was actually modified.

For the analysis of the factors that affect concordance of medication intensification data in narrative and structured sources the study year was calculated as the difference between the year of the provider-patient encounter and the first year of the study (2000). Provider experience with structured EMR medication records was quantified as the number of all structured medication entries that provider had ever made in our institution's EMR before the date of the patient encounter (multiple changes to the same medication record were counted separately).

Treatment Intensification Rate (TIR) was defined as the average number of documented medication intensification events per month of continuously elevated blood pressure. It was calculated as (a) the number of all medication intensifications during the time when the patient had elevated blood pressure divided by (b) the total length (in mo) of all continuous periods where only elevated blood pressure levels were recorded for the patient:

$$\text{TIR} = \frac{(\mathit{Intensifications}_1 + \mathit{Intensifications}_2 + \mathit{Intensifications}_3), N}{(\mathit{HTNPeriod}_1 + \mathit{HTNPeriod}_2 + \mathit{HTNPeriod}_3), \mathit{months}}$$

The length of each individual period of continuously elevated blood pressure was calculated as the difference be-
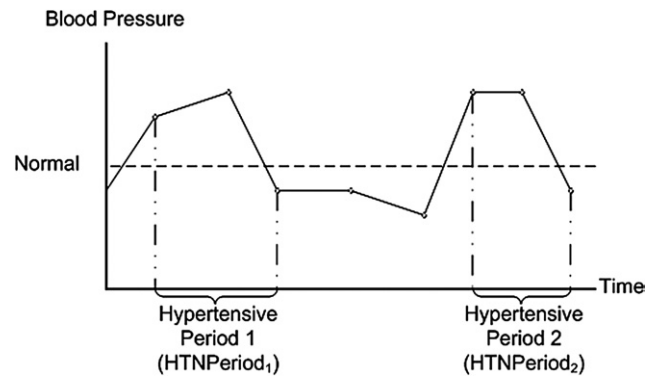

Figure 1. Periods of Continuously Elevated Blood Pressure. Circles represent individual physician–patient encounters.

tween (a) the date when the first elevated blood pressure was documented and (b) the date when the first normal blood pressure was documented (Fig 1). For example, if a patient had elevated blood pressures on 1/1/01, 3/1/01 and a normal blood pressure on 5/1/01, the length of the continuously hypertensive period would be calculated as four months (the difference between 5/1/01 and 1/1/01).

Separate treatment intensification rates were calculated for three data sources: (1) medication intensifications documented in the notes but not structured EMR medication records; (2) medication intensifications documented in the structured EMR medication records but not in the notes; and (3) medication intensifications documented in both structured records and the notes.

Only notes and medication records authored by physicians in primary care practices were used for the analysis. To make the data sets fully comparable, only medication records signed by the physicians who had authored at least one of the study notes were used for the analysis.

Average monthly change in blood pressure (systolic and diastolic) was used to confirm clinical validity of the treatment intensification measures from both sources. Monthly change in blood pressure was computed as the difference between the last and the first blood pressure of the period with continuously elevated blood pressure (see definition above) divided by the length of the period in months:

$$\text{Monthly Change in SBP} = \frac{(SBP_{last} - SBP_{first}), mm\,Hg}{HTNPeriod, months}$$

The last blood pressure used in the calculation was the first blood pressure below the treatment target (if blood pressure eventually normalized) or the last blood pressure recorded during the study period (if blood pressure never normalized). The average monthly blood pressure change for a given patient was calculated as the mean of the monthly blood pressure changes for all periods of continuously elevated blood pressure the patient had during the study.

Statistical Analysis

Summary statistics were constructed using frequencies and proportions for categorical data and using means and standard deviations for continuous variables. Two-sided t test was used for univariate analysis of the difference between the average blood pressure changes over time in patients with different rates of treatment intensification.

To analyze the factors that were associated with concordance of intensifications in structured and narrative data we constructed a hierarchical (multilevel) multiple logistic regression model. We used the GLIMMIX procedure to adjust for clustering within treating physicians and patients.[30,31] The ratio of generalized χ^2 to the degrees of freedom was used to assess goodness of fit of the model.

To determine the relationship between treatment intensification rates and blood pressure changes, we constructed a hierarchical (multilevel) multivariate mixed linear regression model with random effects to account for clustering within treating physicians. Random cluster effects were used to generate correlation structure for intracluster observations as well as account for individual physician effect levels.[32] The model adjusted for the patients' demographic characteristics including age, gender, ethnicity and health insurance status. Visual inspection of residuals plotted against predicted values was used to assess goodness of fit of the model.

The p values were obtained using a type III test for all multivariate analyzes. A type III test evaluates the hypothesis that the covariate significantly improves the model that contains all other covariates.[33] Association significance thresholds were calculated using Simes-Hochberg for multiple testing.[34,35] The SAS version 9.1.3 (SAS Institute, Inc, Cary, NC) was used for all analyses.

Institutional Review Board

Partners HealthCare System institutional review board granted expedited approval of this study and waived the need for informed consent.

Results

Blood Pressure and Anti-hypertensive Medications in Patients with Diabetes

We identified 5,634 patients with documented diagnosis of diabetes, at least one recorded blood pressure level above the recommended target levels and at least two years of follow-up in a primary care clinic during the study period (Table 1). Of these, 4,771 (84.7%) patients had at least one active prescription for anti-hypertensive medications in the EMR by a primary care physician who had seen them at least once during the study period. Study patients had elevated blood pressure documented in the majority (60.9%) of the 85,078 study encounters. On average, their blood pressure was elevated over 61.8% of the time they were being followed during the study period.

Treatment Intensification in Structured and Narrative EMR Data

Anti-hypertensive medication intensifications were documented on 9,819 days in physician notes and on 13,993 days in the structured medication records during the periods of continuously elevated blood pressure. Medication intensifications were documented on the same day in both physician notes and the structured records on 5,627 days (30.9% of all intensifications documented in either source). Of the 8,366 medication intensifications in the structured data that did not have a match in the narrative notes, 5,375 (64.2%) were entered on a day where there was no note by a primary care physician in the EMR.

Table 1 ■ Patient Characteristics

Variable	Value
Study patients, n	5,634
Age*, years (± SD)	64.2 (± 13.5)
Women, n (%)	3,257 (57.8)
Ethnicity	
White, n (%)	3,261 (57.9)
Black, n (%)	1,000 (17.7)
Hispanic, n (%)	915 (16.2)
Other, n (%)	458 (8.1)
English is the primary language, n (%)	4,604 (81.7)
Health insurance*	
Insured, n (%)	3,586 (63.6)
Underinsured†, n (%)	1,878 (33.3)
No prescription coverage‡, n (%)	170 (3.0)
Study follow-up period, years (± SD)	3.9 (± 1.1)
Number of study encounters, n (± SD)	15.1 (± 8.9)
Encounters with elevated blood pressure, n (± SD)	9.1 (± 6.8)
Total intervals of continuously elevated blood pressure, mo (± SD)	29.5 (± 16.2)
Systolic blood pressure§, mm Hg (± SD)	132.0 (± 11.0)
Diastolic blood pressure§, mm Hg (± SD)	75.7 (± 7.2)

*At the end of the study period.
†Includes Medicaid and FreeCare—a program that provides fully or partially (depending on the income) subsidized health care in Massachusetts.
‡Includes Medicare without supplemental insurance and patients with no reported insurance.
§Average across all study encounters with documented blood pressure.

Out of the 5,627 medication intensifications documented on the same date in both sources, the same medication was intensified on 4,987 (88.6%) days. On 743 out of 9,819 (7.6%) days only the class (e.g., "beta-blocker") but not the exact medication being intensified could be determined from computational analysis of physician notes. The medication documented to have been intensified in the structured records belonged to the same class on 196 (23.9%) days. Therefore, the upper bound of the estimate of the number of days when intensification of the same medication was documented in both sources was calculated to be 4,987 + 196 = 5,183 days (92.1% of all days when medication intensification was documented both in the notes and in the structured records).

At the end of the study median age of the 350 primary care physicians who had authored at least one of the study patients' notes was 35; majority were women (Table 2). The number of structured EMR medication records these providers had entered from the first time they used the enterprise EMR until the end of the study (representing their experience with the EMR) ranged from 0 to 76,800. Concordance of intensifications between narrative and structured data were distributed bimodally among the physicians with large peaks around 0% and 50–80% (Fig 2). A smaller peak around 100% was comprised primarily of providers who had fewer than 5 encounters with study patients.

In multivariate analysis the odds of existence of a structured record corresponding to medication intensification documented in the notes increased by 11% for every year of the study (p < 0.0001) and decreased by 19% for every decade of provider's age (p = 0.035). There was a trend for a rise in the

Table 2 ■ Provider Characteristics

Variable	Value
Study providers, n	350
Age*, years (± SD)	39.1 (± 9.8)
Women, n (%)	177 (50.6%)
Fraction of intensifications in narrative documents with corresponding intensifications in structured data, % (± SD)	47.2 (± 30.7)
Structured medication records entered*, n (± SD)	9,175 (± 12,814)

*At the end of the study period.

odds of concordance between intensification in structured and narrative sources with increasing EMR experience (represented by the number of structured medication records ever entered by the provider) which did not reach statistical significance.

Treatment Intensification and Blood Pressure
In univariate analysis patients who had higher rates of medication intensification documented in either structured or narrative data had larger monthly decreases in blood pressure (Fig 3). Patients whose anti-hypertensive medications were intensified at least once a year had 20–50% greater annual decrease in systolic and diastolic blood pressure compared to the patients whose medications were never intensified (p = < 0.01 for all except the relationship between intensifications documented only in the structured notes and change in diastolic blood pressure which did not reach significance).

In multivariate analysis intensifications documented in either narrative or structured data were associated with higher rates of decrease in blood pressure (Table 3). For every additional monthly medication intensification, systolic blood pressure fell by 5–8-mm Hg per month and diastolic blood pressure by 1.5–4-mm Hg per month (p < 0.0001 for all).

Discussion

In this large retrospective cohort study we have compared medication intensification information obtained from narrative and structured data sources in the EMR. Frequency of documented medication intensification were relatively low in both sources −0.060 and 0.076/mo from the notes and structured medication data, respectively—consistent with the previously reported rates of medication intensification for other hypertensive populations.[23,29] While both sources had a similar number of documented intensifications, less than a third of all intensifications were recorded in both structured and narrative data. Concordance between the two information sources increased slightly over the course of the study, possibly reflecting the users' level of comfort and familiarity with the EMR application. However, even by the end of the study less than 40% of all anti-hypertensive medication intensification events were documented in both narrative and structured data.

Several reasons could be contributing to this large discrepancy. Recording all changes to the patient's medications in the narrative notes is required from physicians for billing purposes.[36] At the same time, entering medication information into structured records in the EMR can be time-consuming, particularly for less experienced users. Some physicians may therefore view recording medications in structured lists as duplicative work, leading to avoidance as an expected coping behavior. Even when electronic prescribing is mandatory (as it was in some of the practices during a part of the study period), it is possible for physicians to circumvent the EMR, for example, by calling the pharmacy or instructing the patient to change the medication dose without actually changing the prescription. By contrast, medication changes initiated outside of a face-to-face physician–patient encounter (e.g., by telephone or e-mail) may never be recorded in narrative notes. This possible explanation is supported by the large fraction of medication intensifications in our structured data that were entered on the days when there was no documented physician–patient encounter in the EMR. Anecdotally, as entering prescriptions through EMR rather than on paper has become mandatory in some practices at our institution, compliance has improved. That may have been one of the reasons for a small annual increase in concordance between narrative and structured medication information. However, large discrepancies still remain. Further studies are needed to elucidate the reasons for these discrepancies and identify ways of eliminating them.

Both narrative and structured EMR records have strengths and weaknesses as possible sources of medication information. Structured records may contain more complete information about a particular medication as the users are forced to enter all elements of the prescription. Electronic medication records are also easier to process computationally than narrative text. Consequently a transition towards a greater number of structured data entry options in the EMR has
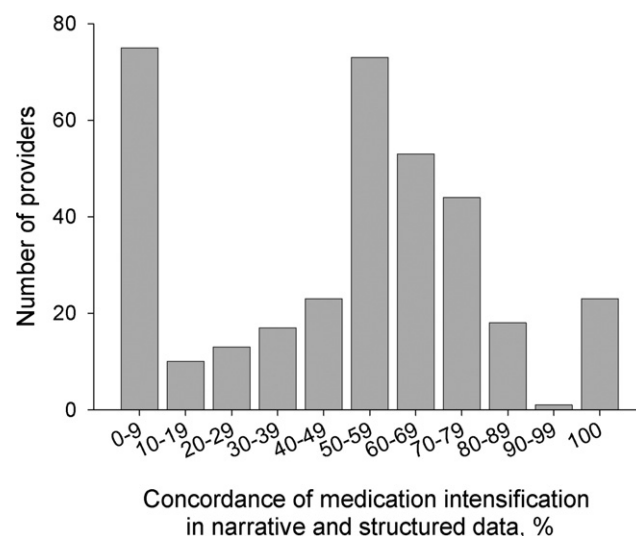

Figure 2. Distribution of Concordance between Medication Intensification Documentation in Narrative and Structured Data among Providers. For each provider who has had an encounter with at least one study patient, the average concordance of medication intensification was calculated as the fraction of anti-hypertensive medication intensifications documented in the notes that had medication intensifications documented in structured EMR records for the same encounter. Intervals of concordance of medication intensification were plotted against the number of providers whose average concordance fell into these intervals.
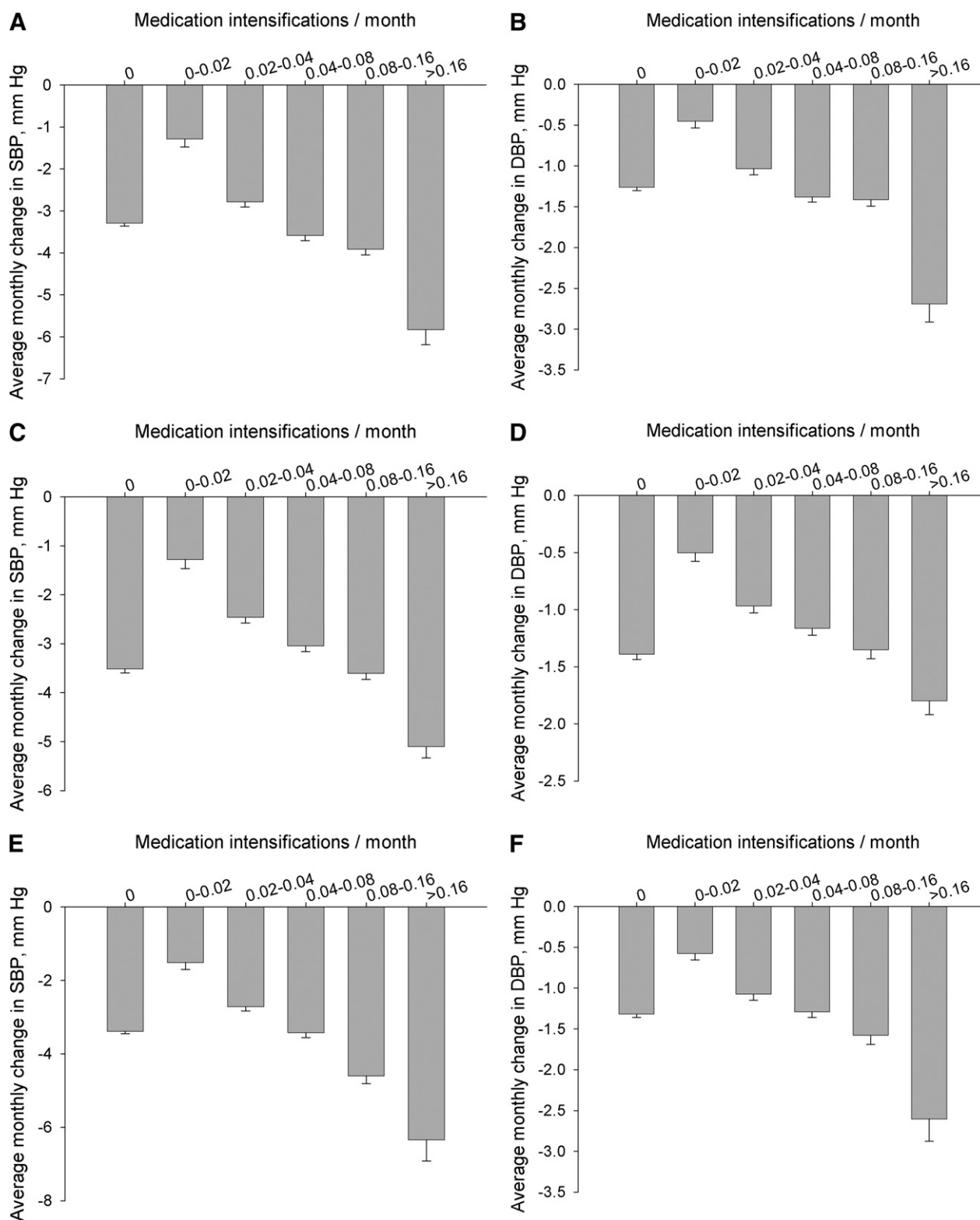
F i g u r e 3.   **A.** Average Monthly Change in Systolic Blood Pressure (SBP) by Frequency of Treatment Intensification Documented Both in the Notes and Structured Records.The average number of anti-hypertensive medication intensifications per month of continuously elevated blood pressure was plotted against the average change in blood pressure, mm Hg/mo. **B.** Average Monthly Change in Diastolic Blood Pressure (DBP) by Frequency of Treatment Intensification Documented Both in the Notes and Structured Records. **C.** Average Monthly Change in Systolic Blood Pressure by Frequency of Treatment Intensification Documented in the Structured Records but Not in the Notes. **D.** Average Monthly Change in Diastolic Blood Pressure by Frequency of Treatment Intensification Documented in the Structured Records but Not in the Notes. **E.** Average Monthly Change in Systolic Blood Pressure by Frequency of Treatment Intensification Documented in the Notes, but Not in the Structured Records. **F.** Average Monthly Change in Diastolic Blood Pressure by Frequency of Treatment Intensification Documented in the Notes, but Not in the Structured Records.

Table 3 ■ Relationship of Anti-Hypertensive Treatment Intensification Documented in Different Sources and Monthly Change in Blood Pressure

Intensification Information Source	Systolic Blood Pressure		Diastolic Blood Pressure	
	Effect[1]	p-value	Effect[1]	p-value
Structured only	−4.7	< 0.0001	−1.5	< 0.0001
Notes only	−8.1	< 0.0001	−2.3	< 0.0001
Both notes and structured[2]	−5.8	< 0.0001	−3.8	< 0.0001

[1]Estimate of a monthly change in blood pressure, mm Hg, associated with increase in treatment intensification rate of one intensification/month.

[2]For intensifications documented both in narrative and structured EMR data.

Effect sizes are the parameter estimates for a mixed linear multiple regression model with the average monthly blood pressure change as the dependent variable and medication intensification rates from all three sources as independent variables.

been advocated.[37] By contrast, narrative text may contain other contextual information that is important for interpretation of the provider action (e.g., about the patient's medication adherence).[38,39] As a result, most EMRs contain a combination of narrative and structured data.[40,41] Furthermore, not all prescriptions may correctly reflect the actual dosing of the medication. For example, it is not uncommon for physicians to prescribe a higher strength of an expensive medication to lower the patient's costs, since the difference in the cost of two different formulations of the same medication is frequently less than the difference in the amount of the chemical ingredient.

Due to the shortcomings of these and other data sources, there is no single gold standard information source for patients' medication information. As a result, concurrent validation (which shows a correlation with an existing test, such as manual chart review[42]) does not provide comprehensive verification of treatment intensification data obtained from either narrative or structured data. To complement this approach we therefore employed predictive validation[42] against the patients' clinical outcomes. It has been well established that anti-hypertensive treatment intensification in real-life clinical environment leads to a decrease in blood pressure levels.[19,23] We were able to demonstrate that this held true for medication intensification information obtained from both narrative and structured EMR records. The relationship remained highly significant in a multivariate analysis that also included patient demographic information and a correction for intra-provider clustering. The deviation from the linear relationship between medication intensification and blood pressure changes observed for encounters with no intensification may have been due to clinical circumstances that rendered intensification inappropriate (e.g., medication nonadherence or acute disease).

To our knowledge, this is the first large-scale analysis that directly compares narrative and structured data in the EMR. It included several thousand patients from two large hospitals that care for patients from all socioeconomic strata. Digital entry of the notes was mandatory for all practices that used the EMR during the study period thus eliminating a potential selection bias. Finally, the data used in our study was validated not only against the original source (which may or may not have been correct itself) but also against the patients' clinical outcomes—a much higher standard that can be difficult to attain.

Our investigation has several limitations. It was restricted in scope to the patients of primary care physicians affiliated with two academic hospitals in Eastern Massachusetts that used an internally developed EMR; this could limit its generalizability. However, although the EMR used in the study was developed internally, its electronic prescribing and narrative note documentation features are similar to many of the commercially available products. The study was restricted to adult patients due to a limited number of pediatric patients with diabetes and hypertension treated at the two study hospitals. Consequently a separate study in children should be carried out to confirm our findings in this population. Patients with short-term (less than 1 mo) elevations of blood pressure were excluded. However, this restriction led to the exclusion of only 59 patients (c 1% of the total number of patients in the study). Therefore, it is unlikely to have significantly affected the study findings. Using individual periods of continuously elevated blood pressure rather than unique patients as the unit of analysis of correlation between treatment intensification and blood pressure may have led to a bias in favor of patients with multiple periods of elevated blood pressure. However, this approach led to a significant reduction in the noise level introduced by normotensive periods and the possible bias was addressed using hierarchical regression models. This retrospective study relied on documentation of relevant findings in the EMR, leading to a possible bias if the documentation was uneven with respect to the study outcomes. Electronic prescribing was not universally mandatory during the study leading to a possible selection bias; however, nearly 85% of the study patients had at least one structured medication record in the EMR. Electronic prescribing became mandatory at some of the practices during the study period which could have affected the study results. The dates when electronic prescribing became mandatory were not available and therefore quantitative analysis of these effects was not feasible. It was not possible to ascertain exactly when patients' blood pressure normalized, limiting the precision of our calculations of the rate of blood pressure change. We considered medication intensifications documented in both sources on the same day identical, even if we could not always establish that they referred to the same medication. However, we were able to establish the equivalence of the medications being intensified in over 90% of the cases. The remaining difference would have tended to bias our multivariate analysis of the relationship of medication intensification information from both sources and the patients' blood pressure towards the null hypothesis by inappropriately decreasing the intensification rate calculated from structured data.

We were not able to obtain pharmacy and/or insurance claims data to complement medication information from EMR records. It is possible that inclusion of these high-validity[43] sources could have helped to resolve some of the discrepancies in the EMR data observed in our study. In the future, point-of-care availability of claims and pharmacy medication records would likely be an important component of closed-loop medication information systems which

would facilitate reconciliation of medication information between different data sources.

Conclusions

Both narrative and structured records in the EMR systems contain valid information about anti-hypertensive medication intensification. Nevertheless significant discrepancies between these two sources are common. Both narrative and structured data should be considered as information sources for research and administrative applications.

References ■

1. Hicks J. The potential of claims data to support the measurement of health care quality. [PhD]. San Diego, CA, RAND; 2003.
2. de Lusignan S, Teasdale S, Little D, et al. Comprehensive computerised primary care records are an essential component of any National Health Information strategy: Report from an international consensus conference. Inform Prim Care 2004; 12(4):255–64.
3. Friedman C, Shagina L, Lussier Y, et al. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc September–Oct 2004;11(5):392–402.
4. Hahn U, Romacker M, Schulz S. MEDSYNDIKATE—A natural language system for the extraction of medical information from findings reports. Int J Med Inform 2002;67(1–3):63–74.
5. Mamlin BW, Heinze DT, McDonald CJ. Automated extraction and normalization of findings from cancer-related free-text radiology reports. AMIA Annu Symp Proc 2003:420–4.
6. Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: Performance evaluation. J Biomed Inform 2006 Dec;39(6):589–99. Epub 2005 Dec 5.
7. Christensen L, HP, Fiszman M, MPLUS. A probabilistic medical language understanding system. Paper presented at: Workshop on Natural Language Processing in the Biomedical Domain, 2002.
8. Hazlehurst B, Frost HR, Sittig DF, et al. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. J Am Med Inform Assoc September–Oct 2005;12(5):517–29.
9. Pakhomov SV, Buntrock J, Chute CG. Prospective recruitment of patients with congestive heart failure using an ad hoc binary classifier. J Biomed Inform Apr 2005;38(2):145–53.
10. Greene RA, Beckman H, Chamberlain J, et al. Increasing adherence to a community-based guideline for acute sinusitis through education, physician profiling, and financial incentives. Am J Manag Care Oct 2004;10(10):670–8.
11. McDonald CJ, Overhage JM, Dexter P, et al. A framework for capturing clinical data sets from computerized sources. Ann Intern Med Oct 15 1997;127(8 Pt 2):675–82.
12. de Lusignan S, van Weel C. The use of routinely collected computer data for research in primary care: Opportunities and challenges. Fam Pract Apr 2006;23(2):253–63.
13. Wolf-Maier K, Cooper RS, Banegas JR, et al. Hypertension prevalence and blood pressure levels in 6 European countries, Canada, and the United States. J Am Med Assoc May 14 2003; 289(18):2363–9.
14. Stratton IM, Kohner EM, Aldington SJ, et al. UKPDS 50: Risk factors for incidence and progression of retinopathy in type II diabetes over 6 years from diagnosis. Diabetologia Febr 2001; 44(2):156–63.
15. Turner RC, Millns H, Neil HA, et al. Risk factors for coronary artery disease in non-insulin dependent diabetes mellitus: United Kingdom prospective diabetes study (UKPDS: 23). BMJ Mar 14 1998;316(7134):823–8.
16. Hovind P, Tarnow L, Rossing P, et al. Predictors for the development of microalbuminuria and macroalbuminuria in patients with type 1 diabetes: Inception cohort study. BMJ May 8 2004;328(7448):1105.
17. Resnick HE, Foster GL, Bardsley J, et al. Achievement of American Diabetes Association clinical practice recommendations among US adults with diabetes, 1999–2002: The national health and nutrition examination survey. Diab Care Mar 2006; 29(3):531–7.
18. Saaddine JB, Engelgau MM, Beckles GL, et al. A diabetes report card for the United States: Quality of care in the 1990s. Ann Intern Med Apr 16 2002;136(8):565–74.
19. Okonofua EC, Simpson KN, Jesri A, et al. Therapeutic inertia is an impediment to achieving the healthy people 2010 blood pressure control goals. Hypertension Mar 2006;47(3):345–51.
20. Phillips LS, Branch WT, Cook CB, et al. Clinical inertia. Ann Intern Med Nov 6 2001;135(9):825–34.
21. Jones DW, Hall JE. Seventh report of the Joint National Committee on Prevention, detection, evaluation, and treatment of high blood pressure and evidence from new hypertension trials. Hypertension Jan 2004;43(1):1–3.
22. Kerr EA, Krein SL, Vijan S, et al. Avoiding pitfalls in chronic disease quality measurement: A case for the next generation of technical quality measures. Am J Manag Care Nov 2001;7(11): 1033–43.
23. Berlowitz DR, Ash AS, Hickey EC, et al. Inadequate management of blood pressure in a hypertensive population. N Engl J Med Dec 31 1998;339(27):1957–63.
24. Turchin A, Kolatkar NS, Grant RW, et al. Using regular expressions to abstract blood pressure and treatment intensification information from the text of physician notes. J Am Med Inform Assoc. Number v-Decis 2006;13(6):691–5.
25. Shah NR, Seger AC, Seger DL, Fiskio JM, Kuperman GJ, Blumenfeld B, Recklet EG, Bates DW, Gandhi TK. Improving acceptance of computerized prescribing alerts in ambulatory care. J Am Med Inform Assoc 2006;13:5–11.
26. Turchin A, Shubina M, Chodos AH, et al. Effect of board certification on antihypertensive treatment intensification in patients with diabetes mellitus. Circulation Feb 5 2008;117(5):623–8.
27. Turchin A, Kohane IS, Pendergrass ML. Identification of patients with diabetes from the text of physician notes in the electronic medical record. Diab Care Jul 2005;28(7):1794–5.
28. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Arch Intern Med Nov 24 1997;157(21):2413–46.
29. Grant RW, Buse JB, Meigs JB. Quality of diabetes care in US academic medical centers: Low rates of medical regimen change. Diab Care Febr 2005;28(2):337–442.
30. Molenberghs G, Verbeke G. Models for Discrete Longitudinal Data, New York; London: Springer, 2005.
31. Witte JS, Greenland S, Kim LL, et al. Multilevel modeling in epidemiology with GLIMMIX. Epidemiology Nov 2000;11(6): 684–8.
32. Greenland S. Principles of multilevel modelling. Int J Epidemiol Febr 2000;29(1):158–67.
33. Searle SR. Linear Models for Unbalanced Data, Hoboken, NJ: Wiley InterScience, 2006.
34. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika 1986;73(3):751–4.
35. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Vol 75. Biometr Trust;1988:800–2.
36. MS. Documentation guidelines for evaluation and management services Available at: http://www.cms.hhs.gov/MLNProducts/downloads/1995dg.pdf Accessed: 12/25/08.
37. Donald CJ. The barriers to electronic medical record systems and how to overcome them. J Am Med Inform Assoc May–Jun 1997;4(3):213–21.
38. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. BMJ May 15 2004;328(7449):1184–7.

39. Berg M, Goorman E. The contextual nature of medical information. Int J Med Inform Dec 1999;56(1–3):51–60.

40. Lovis C, Baud RH, Planche P. Power of expression in the electronic patient record: Structured data or narrative text? Int J Med Inform September 2000;58–59:101–10.

41. Tange HJ, Hasman A, de Vries Robbe PF, et al. Medical narratives in electronic medical records. Int J Med Inform Aug 1997;46(1):7–29.

42. Anastasi A, Urbina S. Psychological Testing, 7th edn, Upper Saddle River, NJ: Prentice-Hall, 1997.

43. Tamblyn R, Lavoie G, Petrella L, et al. The use of prescription claims databases in pharmacoepidemiological research: The accuracy and comprehensiveness of the prescription claims database in Quebec. J Clin Epidemiol Aug 1995;48(8):999–1009.